On the Use of Panel Data in Bayesian Stochastic Frontier Models

Fernández, C.; Osiewalski, J.; Steel, M.F.J.

Publication date:
1996

ON THE USE OF PANEL DATA IN BAYESIAN STOCHASTIC FRONTIER MODELS[1]

Carmen Fernández
CentER and Department of Econometrics, Tilburg University
Tilburg, The Netherlands

Jacek Osiewalski
Department of Econometrics, Academy of Economics
Kraków, Poland

Mark F.J. Steel
CentER and Department of Econometrics, Tilburg University
Tilburg, The Netherlands

ABSTRACT

We consider a Bayesian analysis of the stochastic frontier model with composed error. Under a commonly used class of (partly) noninformative prior distributions, the existence of the posterior distribution and of posterior moments is examined. Viewing this model as a Normal linear regression model with regression parameters corresponding to both the frontier and the inefficiency terms, generates the insights used to derive results in a very wide framework. It is found that in pure cross-section models posterior inference is precluded under this "usual" class of priors. Existence of a well-defined posterior distribution crucially hinges upon the structure imposed on the inefficiency terms. Exploiting panel data naturally suggests the use of more structured models, where Bayesian inference can be conducted.

Keywords: Composed error; Existence of posterior; Identification; Improper priors; Posterior moments.

JEL classification: C11; C23; D24.

[1] Corresponding author: Carmen Fernández, Department of Econometrics, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands. Phone: +31/13/466 3254. Fax: +31/13/466 3280. E-mail: carmen@kub.nl.

1. INTRODUCTION

Since its introduction in Aigner, Lovell and Schmidt (1977) and Meeusen and van den Broeck (1977), the stochastic frontier model with composed error has proved a very useful vehicle for the analysis of production or cost frontiers when economic agents need not necessarily reach the frontier. The distance with respect to the frontier is then interpreted as technical or cost inefficiency. This leads to a one-sided deviation, which is combined with the "usual" measurement error, reflecting that the frontier is not known, but needs to be inferred from the data. Whereas the measurement error is typically chosen to be Normally distributed, the inefficiency term has been allocated various types of one-sided distributions in the literature.

Classical analyses of this model can be found in *e.g.* Meeusen and van den Broeck (1977) with an Exponential inefficiency distribution, Aigner, Lovell and Schmidt (1977) (half-Normal), Stevenson (1980) (truncated Normal) and Greene (1990) (Gamma). Applications of this model on cross-sectional data abound and are usually conducted in the framework of one of the references cited above. A Bayesian analysis of this model was introduced for cross-sectional data in van den Broeck, Koop, Osiewalski and Steel (1994) under a variety of inefficiency distributions.

Panel or longitudinal data provide us with multiple observations over time for each economic agent (generically denoted by "firms"). The use of panel data in stochastic frontier analysis was pioneered by Pitt and Lee (1981) and Schmidt and Sickles (1984). Whereas Pitt and Lee (1981) exclusively present a Maximum Likelihood analysis, Schmidt and Sickles (1984) remark that the assumption of constant efficiencies over time allows us to use a "within" estimator in a fixed effects context. Unlike Maximum Likelihood, this does not require a distributional assumption on the inefficiency terms. Inference can then only be conducted on relative, rather than absolute efficiencies, however. Koop, Osiewalski and Steel (1996) present a Bayesian interpretation of this fixed effects procedure, and point out some implicit assumptions and consequences in the context of a hospital cost frontier.

Most noninformative or reference prior distributions used in Bayesian analyses are improper, *i.e.* they do not integrate to a finite number over the parameter space. This is the case for the prior distributions used in *e.g.* van den Broeck, Koop, Osiewalski and Steel (1994) and Koop, Osiewalski and Steel (1996). Whereas an improper prior can, in many cases, be transformed into a well-defined, proper posterior distribution, the existence of such a posterior is not guaranteed [see *e.g.* O'Hagan (1994, p.79)] and needs to be verified in each particular situation. Let us now explain these concepts in some detail.[2] It is well-known that a sampling probability density function (p.d.f) $p(y|\theta)$ for $y \in \Re^\mu$ and a prior p.d.f. $p(\theta)$ for a parameter $\theta \in \Re^\nu$ uniquely define a joint probability distribution on $\Re^\mu \times \Re^\nu$, with density

$$p(y,\theta) = p(y|\theta)p(\theta). \tag{1.1}$$

The essence of the Bayesian paradigm lies in the dual decomposition of this joint distribu-

[2] For the sake of simplicity, we shall reason in terms of density functions, but the essential argument in no way hinges upon this.

tion into the marginal (predictive) distribution of y, with density

$$p(y) = \int_{\Re^{\nu}} p(y|\theta)p(\theta)d\theta \tag{1.2}$$

and the conditional (posterior) distribution of θ given y, characterized by the p.d.f.

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}. \tag{1.3}$$

Bayesian prediction is based on $p(y)$, whereas posterior inference is conducted through $p(\theta|y)$. This decomposition relies upon the rules of probability calculus, and, thus, whenever the prior $p(\theta)$ is improper the question whether Bayesian inference is still possible immediately arises. In this case, we can distinguish two possible situations:

Situation (A): the predictive distribution is σ-finite, *i.e.* $p(y) < \infty$ for all $y \in \Re^{\mu}$, barring a set of Lebesgue measure zero.
Situation (B): the predictive distribution is not σ-finite, *i.e.* there exists a set of y's of positive Lebesgue measure in $\Re^{\mu}$ for which $p(y) = \infty$.

Mouchart (1976) shows that only Situation (A) still allows for a decomposition of the joint distribution into a σ-finite marginal distribution for y (predictive) and a conditional probability distribution of θ given y (posterior). The corresponding density functions are again given by (1.1) – (1.3). In Situation (B) such a decomposition is not possible. This, however, does not exclude the possibility that $p(y) < \infty$ for certain samples y; for such values of y the expression in (1.3) is still a p.d.f. on the parameter space, which could be associated with a "posterior distribution". Nevertheless, it must be realized that such a construction does not have the interpretation of the conditional probability distribution of θ given y, and is, thus, of questionable use for conducting Bayesian inference. Throughout the paper, Situation (B) will be referred to as "lack of existence of the posterior distribution"; alternatively, we shall say that the posterior distribution is not "well-defined" in this case. If, in addition, $p(y) = \infty$ for all $y \in \Re^{\mu}$, we shall explicit this in the text. Once the issue of existence of the posterior distribution is settled, we have to consider that posterior moments of certain quantities of interest can also fail to exist, even if the prior is proper.

Since the required calculations in empirical applications are typically done by high-dimensional numerical integration[3] on a computer, the actual results may not reflect the lack of existence of the posterior distribution or moments. In order to have valid answers, however, we do need to make sure that the posterior is well-defined and the relevant posterior moments all exist. This paper addresses these issues in detail for the stochastic frontier model with composed error.

The availability of panel data often leads to the possibility to reasonably impose some structure on the inefficiency terms. The present paper shows that, under the "usual" class of (partly) noninformative priors, such structure is essential for the existence of the posterior distribution. In itself, a total lack of structure in the vector of inefficiencies can

[3] See Osiewalski and Steel (1996) for a survey of numerical tools in this context.

preclude Bayesian analysis under this class of priors. Changing the specification of the frontier, the prior on the frontier parameters or introducing probabilistic links between inefficiency terms can not solve the problem.

These results emerge from considering the stochastic frontier model as a simple Normal linear regression model where the regression parameters are both the frontier parameters and the structured vector of inefficiencies. In this way, the fact that a lack of structure on inefficiencies basically leads to an overparameterized location transpires, and the posterior can be shown not to exist. Previously, the stochastic frontier model was often interpreted as a location mixture of Normals [see also Geweke (1995)], where the inefficiency terms were first integrated out with their assigned distribution, resulting in the marginalized likelihood. This interpretation is more natural in a classical random effects framework, where the inefficiency distribution is considered part of the sampling distribution and inference is based on this marginalized likelihood. The latter is typically very complicated [see *e.g.* Stevenson (1980) and Greene (1990)] and defies intuitive understanding. Inferences on efficiencies are, of course, of interest, but they are then often conducted in a second step using the Jondrow, Lovell, Materov and Schmidt (1982) approach. Even though the marginalized likelihood also formed the basis of the Monte Carlo Importance Sampling approach in van den Broeck, Koop, Osiewalski and Steel (1994) and the Bayesian results in Ritter (1993), the inefficiency terms and the frontier parameters are formally on equal footing for Bayesians. This is all the more relevant since we are interested in both the frontier parameters and the inefficiencies. Viewing the model as a regression model with a prior distribution on both inefficiencies and the other parameters creates a lot of additional insight and essentially renders the derivation of our results feasible in a wide context.

In addition, this way of considering the model underlines a frequently occurring identification problem. The issues of identification and existence of the posterior distribution are logically separate. As a result of the prior structure commonly assumed for this model, this fundamental difference is clearly illustrated in this context.

Section 2 introduces the model and briefly discusses identification. We then present the main results on existence of the posterior distribution (Section 3) and moments (Section 4). A final section concludes and reinterprets previously published results in the light of our findings.

2. THE MODEL

The linear stochastic frontier model with composed error takes the following general form

$$y = X\beta - \gamma + \tau^{-1}v, \tag{2.1}$$

where y is a $TN \times 1$ vector grouping T observations of the logarithm of output (or the negative of log cost) for N firms. X denotes the corresponding $TN \times k$ matrix of exogenous regressors describing the frontier, with $\beta = (\beta_1, \dots, \beta_k)' \in \mathcal{B} \subseteq \Re^k$ as a vector of regression coefficients; often, theoretical considerations will lead to regularity conditions on β, which will restrict the parameter space $\mathcal{B}$ to an open subset of $\Re^k$, still k-dimensional and possibly depending on X. Measurement error is modelled through the vector v, which has a standard TN-dimensional Normal distribution, and the precision parameter τ lies in $\Re_+$. The distinguishing feature of the stochastic frontier model in (2.1) is the composed error structure, which is expressed through the term $\gamma \in \Re_+^{TN}$; this reflects the fact that firms need not lie on the frontier, but can display inefficiencies, i.e. one-sided departures from the frontier. As usual, the conditional distribution of v given (β, τ, γ) is assumed not to depend on the latter vector.

We adopt the convention of ordering the observations such that $y' = (y_1', \ldots, y_N')$, where $y_i' = (y_{(i,1)}, \ldots, y_{(i,T)})$ contains the T observations for firm i. In general, the double subscript (i,t) corresponds to the i^{th} firm at time t.

Of particular interest in this context will be the efficiencies, defined as $\exp(-\gamma_{(i,t)}) \in (0,1)$. In practice, models will differ in the amount of structure assumed on γ. In order to accommodate this, we shall express γ in terms of an M-dimensional vector z ($M \leq TN$):

$$\gamma = Dz, \tag{2.2}$$

where D is a known or exogenous $TN \times M$ matrix and $z \in \mathcal{Z}$ with $\mathcal{Z} = \{z = (z_1, \ldots, z_M)' \in \Re^M : Dz \in \Re_+^{TN}\}$. We index the components of z with one single subscript since they do not necessarily correspond to a unique firm and time period. Note that $\mathcal{Z}$ depends on D. As important examples we can mention the following cases:

Case (i). *$D = I_{TN}$, i.e. $M = TN$ and $\gamma = z$. This case implies that each firm in each period has its specific inefficiency term $\gamma_{(i,t)} = z_{(i-1)T+t}$. No further structure is imposed and the possible panel character of the data is not exploited. This model corresponds to Model II in Pitt and Lee (1981).*

Case (ii). *$D = I_N \otimes \iota_T$, where ι_T is a T-dimensional vector of ones and $\otimes$ denotes the Kronecker product. This means that $M = N$ and $\gamma_{(i,t)} = z_i$; thus, z_i denotes the inefficiency of firm i, assumed constant over time. In this "individual effects" model we clearly use the panel structure of the data, as in e.g. Pitt and Lee (1981, Model I), Schmidt and Sickles (1984) and Koop, Osiewalski and Steel (1996). Trivially, in a pure cross-section setup (T=1) as in Greene (1990) and van den Broeck, Koop, Osiewalski and Steel (1994), Cases (i) and (ii) coincide.*

Case (iii). *$D = S \otimes I_T$, where S is a matrix of dimension $N \times J$ with $J \leq N$. Then $M = TJ$ and the inefficiency of firm i at time t is given by $\gamma_{(i,t)} = s_i'\tilde{z}_t$, where s_i' denotes the i^{th} row of S and $\tilde{z}_t = (z_t, z_{T+t}, \ldots, z_{(J-1)T+t})'$. Let us mention two possible implementations.*
<u>Case (iiia):</u> Assuming that firms belong to one of J clusters characterized by the same efficiencies (still varying over time) corresponds to taking S a selection matrix, where the $(i,j)^{th}$ element takes the value one if firm i belongs to cluster j and zero otherwise. $\tilde{z}_t$ is then the vector of inefficiencies of the J clusters at time t. Clearly, for $J = N$ we are back in Case (i).
<u>Case (iiib):</u> A second possible choice for S is a matrix of regressors, where s_i' is a J-dimensional vector of time-invariant characteristics of the i^{th} firm. In this case, $\tilde{z}_t$ groups the regression coefficients corresponding to time t.

Case (iv). *$D = I_N \otimes R$ with R a matrix of dimension $T \times H$, where $H \leq T$. Thus, $M = NH$ and $\gamma_{(i,t)} = r_t'z_i^*$, where r_t' denotes the t^{th} row of R and $z_i^* = (z_{(i-1)H+1}, \ldots, z_{iH})'$.*

We distinguish the following cases, analogously to Case (iii):
<u>Case (iva):</u> R is a selection matrix with the $(t,h)^{th}$ element equal to unity if period t belongs to cluster h and zero otherwise. In this case, we allow inefficiencies to vary over firms, but assume that they remain constant in each cluster of (usually contiguous) time periods. Clearly, z_i^* groups the inefficiencies of firm i over the H time clusters. If $H = 1$ we are back in Case (ii) and for $H = T$ this case reduces to Case (i).
<u>Case (ivb):</u> Alternatively, R could group H time-varying regressors common to all firms. As an example, we could assume that firm-specific inefficiencies vary as a polynomial function of time, and choose $r_t' = (1, t, \ldots t^{H-1})$, as introduced in Cornwell, Schmidt and Sickles (1990) for $H = 3$. Kumbhakar (1990) proposes a parametric function of time for r_t' with $H = 1$.

Combining (2.1) with (2.2) leads to the following expression for the sampling density

$$p(y|\beta, \tau, z) = f_N^{TN}(y|X\beta - Dz, \tau^{-2} I_{TN}), \tag{2.3}$$

i.e. the p.d.f. of a TN-dimensional Normal distribution with mean $X\beta - Dz$ and covariance matrix $\tau^{-2} I_{TN}$. Throughout the paper, conditioning on X and D will not be stated explicitly.

We shall assume that the prior for the parameters in (2.3) has the following density function on $\mathcal{B} \times \Re_+ \times \mathcal{Z}$:

$$p(\beta, \tau, z) \propto \tau^{-1} p(z), \tag{2.4}$$

where $p(z)$ can be either proper or improper. The prior density of (β, τ), $p(\beta, \tau) \propto \tau^{-1}$, corresponds to the usual non-informative prior for the linear regression model, but with the parameter space of β possibly restricted to the regularity region. This prior choice has been typically made in the Bayesian stochastic frontier literature [see *e.g.* van den Broeck, Koop, Osiewalski and Steel (1994) and Koop, Osiewalski and Steel (1995, 1996)]. Assuming a distribution for z, however, is not specific to the Bayesian treatment of stochastic frontiers. In classical random effects models a distribution on z is introduced as part of the sampling model, usually indexed by some parameter. In Cases (i), (ii), (iiia) and (iva), the parameter space of z will be $\mathcal{Z} = \Re_+^M$. For such $\mathcal{Z}$ many different distributions have been proposed in the classical literature: half-Normal [Aigner, Lovell and Schmidt (1977)], truncated Normal [Stevenson (1980)], Exponential [Meeusen and van den Broeck (1977)] and Gamma [Greene (1990)]. In a Bayesian framework, $p(z)$ is usually elicited in a hierarchical fashion, leading to *e.g.* marginal distributions of the inverted Beta type[4] as in van den Broeck, Koop, Osiewalski and Steel (1994) and Koop, Osiewalski and Steel (1996). The VED model of Koop, Osiewalski and Steel (1996) introduces a dependence of $p(z)$ on firm characteristics. We remark that this way of allowing for efficiencies to vary with properties of the firms does not introduce exact links between the inefficiency terms as in Case (iiib): the dependence is now implemented in a probabilistic fashion and the dimension of $\mathcal{Z}$ is not reduced.

Classical maximum likelihood analyses of the random effects model as well as the Bayesian analyses of Ritter (1993) and van den Broeck, Koop, Osiewalski and Steel (1994)

[4] Also called Beta prime distribution. See (A.67) in Zellner (1971, p.376) for the probability density function and properties of the distribution.

are based on the marginal likelihood after integrating out z. This results in a usually very complicated expression which leaves little scope for understanding the structure of the model. In contrast, our analysis here is conducted on the sampling model in (2.3) [complemented with the prior in (2.4)], which has a very transparent Normal linear regression structure.

In the context of the regression model in (2.3), the issue of identification becomes especially relevant. Even though we shall assume throughout that X is of full column rank k, the column rank of the entire design matrix $(X : D)$ is often deficient. Typically, the frontier requires an intercept, say β_1, *i.e.* ι_{TN} is the first column of X. Furthermore, the choices of D in Cases (i), (ii), (iiia) and (iva) will imply that $D\iota_M = \iota_{TN}$, so that $(X : D)$ will not be of full column rank, but typically of rank $k + M - 1$. In the latter case, regarding the parameters β_1 and z, the sampling model only provides information on $(\beta_1, z')a$ when $\iota'_{M+1}a = 0$. Other linear functions of β_1 and z are simply not identified. Consider the reparameterization of $(\beta_1, z')'$ to $(\tilde{\phi}'_1, \phi_2)'$ where $\tilde{\phi}_1 = A(\beta_1, z')'$ with A an $M \times (M+1)$ matrix of full row rank such that $A\iota_{M+1} = 0$ and ϕ_2 is a scalar that completes the transformation. Since the likelihood only depends on $\phi_1 = (\tilde{\phi}'_1, \beta_2, \ldots, \beta_k, \tau)'$, there is no updating of ϕ_2 given ϕ_1: the data are "conditionally uninformative for ϕ_2 given ϕ_1" in the terminology of Poirier (1995). From a Bayesian perspective, there are two possibilities. If $p(\beta, \tau, z)$ is such that the prior of ϕ_2 given ϕ_1 is not well-defined,[5] then the posterior distribution does not exist and no Bayesian inference can be conducted. In that case, we could reduce the dimension of $(\beta_1, z')'$ from $M + 1$ to M and conduct inference on relative rather than on absolute efficiencies [see the SIE model of Koop, Osiewalski and Steel (1996)]. If, on the other hand, the implied prior of ϕ_2 given ϕ_1 is well-defined, a Bayesian analysis is not excluded and can be conducted if the marginal posterior $p(\phi_1|y)$ is also proper. In that case, prior dependence between ϕ_1 and ϕ_2 will be required to marginally update ϕ_2. Inference on absolute efficiencies is then possible, despite the deficient column rank of $(X : D)$.

3. EXISTENCE OF THE POSTERIOR DISTRIBUTION

In this Section we shall investigate conditions under which the model in (2.3) – (2.4) allows for a posterior distribution, and thus can be used for Bayesian inference. We remind the reader that the prior distribution of (β, τ) in (2.4) is not proper and we do not necessarily assume a proper prior for z either. As a consequence, the question of existence of the posterior distribution arises, as explained in the Introduction.

We shall implicitly assume throughout the paper that

$$r(X) = k,$$

where $r(\cdot)$ denotes the rank. Our results for the Bayesian model (2.3) – (2.4) are summarized in the following Theorems.

[5] *i.e.* the joint distribution of (ϕ_1, ϕ_2) can not be decomposed into a σ-finite marginal distribution of ϕ_1 and a conditional probability distribution of ϕ_2 given ϕ_1; this corresponds to Situation (B) in the Introduction applied to this prior decomposition.

Theorem 1. *Consider the entire $TN \times (k+M)$-dimensional design matrix $(X : D)$ in (2.3).*
(i) If $r(X : D) < TN$, then the posterior distribution exists for any proper $p(z)$;
(ii) if $r(X : D) = TN$, then the posterior does not exist.

Proof: see Appendix. •

Theorem 1 stresses the importance of the structural specification of the model through the rank of the matrix of all regressors in the linear regression model (2.3). If the latter rank is TN, which holds *e.g.* in Case (i), no prior density function on z (proper or improper) can lead to a posterior distribution. Indeed, it is shown in the Appendix that $p(y) = \infty$ for any sample y in the set

$$W = \{y \in \Re^{TN} : y = X\beta - Dz \text{ for } \beta \in \mathcal{B} \text{ and } z \in C\}, \tag{3.1}$$

where

$$C = \{z \in \mathcal{Z} : \text{ for all } z^* \text{ in some neighbourhood of } z, p(z^*) \geq K > 0 \text{ for some constant } K\}. \tag{3.2}$$

If $r(X : D) = TN$, then W has positive Lebesgue measure in $\Re^{TN}$, which implies that the predictive distribution is not σ-finite. A widely-used model in practice corresponds to Case (i) with an intercept taking values in all of $\Re$ included in the frontier [see *e.g.* van den Broeck, Koop, Osiewalski and Steel (1994) and Koop, Steel and Osiewalski (1995)]. Then, if $p(z)$ is continuous and strictly positive for all $z \in \mathcal{Z} = \Re_{+}^{TN}$ the set W becomes all of $\Re^{TN}$ and, therefore, $p(y) = \infty$ for any sample y. We are thus in an extreme case of Situation (B) in the Introduction and Bayesian inference is certainly precluded. For other specifications of X and D such that $r(X : D) = TN$, W does not necessarily cover all of $\Re^{TN}$ and $p(y)$ can be finite for some samples. However, for the reasons explained in the Introduction, we would hesitate to conduct Bayesian inference in this case even if $p(y) < \infty$ for the observed sample.

In summary, in the pure cross-section model or if the panel structure is not used, the stochastic frontier model in (2.3) does not lead to a posterior distribution under the prior in (2.4). If, on the other hand, the rank of $(X : D)$ is smaller than TN, Theorem 1 (i) tells us that a proper prior on z suffices for a well-defined posterior distribution; clearly, this situation always arises when the number of columns of $(X : D)$, $k + M$, is smaller than TN. In a panel context with $T > 1$, the latter condition translates into $k < (T-1)N$ for Case (ii), $J < N - (k/T)$ for Case (iii) and $H < T - (k/N)$ for Case (iv). If we only have cross-section data ($T = 1$), the structure of Case (iii) will still lead to posterior inference if the number of firm clusters or firm characteristics $J < N - k$.

In most cases of practical interest the prior of z will be given through a proper density function. In this case, existence of the posterior distribution is entirely determined by the rank of $(X : D)$ and the actual form of the prior of z is irrelevant. This is the object of the following Corollary:

Corollary 1. *If $p(z)$ is a probability density function, the posterior distribution is well-defined if and only if $r(X : D) < TN$.*

Proof: immediate from Theorem 1. •

We can also find priors $p(z)$ such that the existence of the posterior distribution is excluded even when $r(X : D) < TN$. Theorem 2 provides a sufficient condition for such a situation:

Theorem 2. *If $p(z)$ is such that $\int_{\mathcal{Z}_0} p(z)dz = \infty$ for some bounded set $\mathcal{Z}_0 \subset \mathcal{Z}$, then the posterior distribution does not exist. Furthermore, $p(y) = \infty$ for all $y \in \Re^{TN}$.*

Proof: see Appendix. •

The following example illustrates this result:

Example 1. *Consider $\mathcal{Z} = \Re_+^M$ [e.g. Cases (i), (ii), (iiia) and (iva)] and the hierarchical prior where z_l, $l = 1, \ldots, M$, are independently Gamma distributed given a parameter $\lambda \in \Re_+$ with*

$$p(z_l|\lambda) = \{\Gamma(\alpha)\}^{-1} \lambda^{\alpha} z_l^{\alpha-1} \exp(-\lambda z_l)$$

with $\alpha > 0$ fixed and $p(\lambda) \propto \lambda^{-1}$. Since we put the "usual" noninformative prior on the precision parameter λ, this may seem interesting as a way to reflect a lack of prior information. However, it can easily be shown that the implied marginal prior density $p(z)$ of $z = (z_1, \ldots, z_M)'$ is nonintegrable in a bounded neighbourhood of the origin; thus, by applying Theorem 2, we know that $p(y) = \infty$ for all y in $\Re^{TN}$ and, therefore, there is no posterior distribution. •

The lack of existence of the posterior mentioned in Example 1 was noted in Ritter (1993). However, he considers only Case (i), so that from Theorem 1 (ii) we know that no prior density on z would lead to a posterior distribution in this model.

For the models where $r(X : D) < TN$, we have, so far, considered proper priors on z [Theorem 1 (i)] and priors that attach infinite mass to a bounded set (Theorem 2). Let us now examine the intermediate case where $p(z)$ need not be proper but is a bounded function.

Theorem 3. *If $r(X : D) = k + M < TN$, then the posterior distribution exists for any bounded prior density $p(z)$.*

Proof: Since $p(z)$ is bounded by some constant, we can simply focus on the Normal linear regression model in (2.3) with the prior density $p(\beta, \tau, z) \propto \tau^{-1}$, which leads to a proper posterior under the assumption of Theorem 3 [see Zellner (1971, Ch.III)]. •

As discussed in Section 2, many practical contexts will exhibit an identification problem, which precludes the application of Theorem 3.

In summary, we stress that under the prior in (2.4), the posterior distribution is not well-defined in Case (i). As a general recommendation to practitioners, we would thus urge to avoid Case (i) by either using the possible panel structure of the data [through clustering over periods, as in Cases (ii) and (iva) or deterministic links between time periods, as in Case (ivb)], or by clustering over or linking firms [as in Cases (iiia) and (iiib)].

Should we insist on using Case (i) (or any other structure that would imply that $r(X : D) = TN$), then we need to change the prior structure on (β, τ). The following Proposition collects some relevant results.

Proposition 1. *If in the Bayesian model (2.3) – (2.4) with $r(X : D) = TN$ we replace the prior (2.4) by*
(i)

$$p(\beta, \tau, z) \propto \tau^{n_0 - 1} \exp(-\tau^2 a_0) p(z) \tag{3.3}$$

with $n_0 \geq 0$ and $a_0 > 0$, then the posterior will exist for any proper $p(z)$, provided $k < TN + n_0$;
(ii)

$$p(\beta, \tau, z) \propto \tau^{-1} p(\beta, z) \tag{3.4}$$

with any density function $p(\beta, z)$ (proper or improper) on $\mathcal{B} \times \mathcal{Z}$, the posterior will not exist.

Proof: simple modifications of the proofs of Theorem 1 (i) and (ii) respectively lead to Proposition 1 (i) and (ii). •

When $n_0 > 0$, (3.3) corresponds to a Gamma distribution on τ^2, but even for the limiting case $n_0 = 0$ the posterior will exist. Note that the functional form of (3.3) tends to (2.4) as n_0 and a_0 both approach zero. In practice, we shall reflect an absence of strong prior information through choosing small values of n_0 and a_0. As a guideline, choosing $n_0 = 0$ and $a_0 = 10^{-6}$ the ratio between (3.3) and (2.4) is in the range $(0.99, 1)$ for values of τ smaller than 100.

In sharp contrast to the result in Proposition 1 (i), changing (2.4) to (3.4) will not solve the problem of lack of a posterior distribution. To put this asymmetric behaviour in perspective, we note that whenever $r(X : D) = TN$, the mean $X\beta - Dz$ in (2.3) lies in a set of positive Lebesgue measure in $\Re^{TN}$. Thus, the location has the same dimension as the vector of observables y, and Proposition 2 of Fernández, Osiewalski and Steel (1995a) can be shown to imply that the posterior distribution does not exist under the prior in (3.4). Effectively, we have overparameterized the location of the model. Thus, changing the prior density of (β, z) will not resolve the problem.

If we are interested in flexible functional form or seminonparametric specifications for the frontier, we can consider adding a nonlinear function $f(\cdot)$ of some parameter α to the frontier to obtain

$$y = X\beta - Dz + f(\alpha) + \tau^{-1} v,$$

as *e.g.* in Koop, Osiewalski and Steel (1994). The next Proposition states that in such a model the posterior does not exist either whenever $r(X : D) = TN$.

Proposition 2. *Consider the sampling density $p(y|\beta, \tau, z, \alpha) = f_N^{TN}(y|X\beta - Dz + f(\alpha),$ $\tau^{-2} I_{TN})$, where $f(\alpha)$ is a continuous function of some q-dimensional parameter α, and the prior $p(\beta, \tau, z, \alpha) \propto \tau^{-1} p(\beta, z, \alpha)$, with $p(\beta, z, \alpha)$ proper or improper. Then, if $r(X : D) = TN$, the posterior distribution does not exist.*

Proof: see Appendix. •

A leading example where $r(X : D) = TN$ is Case (i), with $D = I_{TN}$. From Proposition 2, the lack of existence of the posterior immediately follows, irrespective of the functional form of the frontier. If, in addition to $D = I_{TN}$, the frontier is assumed to have an intercept, and the prior density $p(\beta, z, \alpha)$ is continuous and strictly positive in $\mathcal{B} \times \mathcal{Z} \times \mathcal{A}$, where $\mathcal{A} \subseteq \Re^q$ has positive Lebesgue measure, then the proof of Proposition 2 implies that $p(y) = \infty$ for all samples $y \in \Re^{TN}$.

Finally, the assumption of Normality is not crucial at all for these results of nonexistence of the posterior distribution. From Proposition 2 of Fernández, Osiewalski and Steel (1995a) we can deduce that they apply to a very wide class of multivariate distributions for v, such as any spherical distribution or any v-spherical distribution [see Fernández, Osiewalski and Steel (1995b)] with the isodensity contours bounded away from the origin.

4. EXISTENCE OF POSTERIOR MOMENTS

Let us now examine whether posterior moments of quantities of interest are finite. All results in this Section will relate to the Bayesian model in (2.3) – (2.4) and, as indicated in Section 3, $r(X) = k$. We shall assume throughout this Section that

$$r(X : D) < TN,$$

since we know from Theorem 1 (ii) that otherwise the posterior distribution does not exist.

The results about the existence of the posterior distribution derived in the previous Section all hold for any set $\mathcal{B}$ of positive Lebesgue measure in $\Re^k$. We remind the reader that $\mathcal{B}$ is usually obtained through imposing certain regularity conditions on the frontier, and can take many different forms. The particular choice of $\mathcal{B}$ could crucially influence the existence of posterior moments; as an example, if $\mathcal{B}$ is a bounded set all posterior moments of β of positive order will clearly exist, whereas this does not generally hold if $\mathcal{B}$ is unbounded. The results in this Section correspond to $\mathcal{B} = \Re^k$. Of course, if a posterior moment exists for $\mathcal{B} = \Re^k$, it will also exist for any other choice of $\mathcal{B} \subset \Re^k$; thus, all results in this Section that guarantee existence of moments also hold under any other $\mathcal{B}$.

As mentioned in Section 2, the primary focus of the analysis of stochastic frontier models with composed error are the firm efficiencies $\exp\{-\gamma_{(i,t)}\}$.

Theorem 4. *If the posterior distribution exists, all marginal or product moments of positive order of the efficiencies are finite, i.e.*

$$E\left(\prod_{\substack{i=1,\dots,N \\ t=1,\dots,T}} \exp\{-m_{(i,t)}\gamma_{(i,t)}\}|y \right) < \infty$$

for any $m_{(i,t)} \geq 0$.

Proof: immediate as $\exp\{-\gamma_{(i,t)}\} \in (0,1)$. •

Theorem 4 covers marginal moments, but also product moments across firms or time periods.

For the precision parameter τ we consider moments of both positive and negative orders, since negative moments of τ are the corresponding positive moments of the scale parameter $\sigma = \tau^{-1}$.

Theorem 5. *If $p(z)$ is proper, then*

$$E(\tau^m|y) < \infty \quad \text{if and only if} \quad m > -(TN-k).$$

Proof: a very slight modification of the proof of Theorem 1 (i). •

Results for the moments of the frontier coefficients β are less straightforward to derive. The remainder of this Section will focus on marginal moments of the components of β of positive order. Using the inequality

$$\prod_{l=1}^{k} |\beta_l|^{m_l} \leq \sum_{l=1}^{k} \frac{m_l}{m} |\beta_l|^m, \qquad \text{where } m_l \geq 0 \text{ and } m = \sum_{l=1}^{k} m_l > 0$$

[see *e.g.* Magnus and Neudecker (1988, p.202)], we can assure that if all m^{th} order marginal moments are finite, the $(m_1, \ldots, m_k)^{th}$ order product moment will exist for any nonnegative $m_1, \ldots, m_k$ that sum up to m.

Our main results concerning marginal posterior moments for β are summarized in the next three Theorems:

Theorem 6.
(i) If $m \geq TN - k$ then $E(\beta_l^m|y)$ does not exist;
(ii) for $0 \leq m < TN - k$ we obtain that $E(\beta_l^m|y)$ exists if and only if the following two conditions hold:

$$\int_{\mathcal{Z}} \{(y + Dz)' \, M_X \, (y + Dz)\}^{-(TN-k-m)/2} p(z) dz < \infty \tag{4.1}$$

and

$$\int_{\mathcal{Z}} |w_l'(y + Dz)|^m \{(y + Dz)' \, M_X \, (y + Dz)\}^{-(TN-k)/2} p(z) dz < \infty, \tag{4.2}$$

where w_l' is the l^{th} row of $(X'X)^{-1}X'$ and $M_X = I_{TN} - X(X'X)^{-1}X'$.

Proof: see Appendix. •

Theorem 6 (i) tells us that marginal posterior moments of β_l of order $TN - k$ or higher will never exist, as is the case for σ from Theorem 5. For orders smaller than $TN - k$, Theorem 6 (ii) provides a full characterization of existence of posterior moments. It is shown in the Appendix [see $(A.3)$] that when $r(X : D) < TN$, $(y + Dz)'M_X(y + Dz) \geq K(y)$ for some function $K(y)$ which is strictly positive for all $y \in \Re^{TN}$ except for a set of Lebesgue measure zero. If $p(z)$ is proper, (4.1) is then immediately fulfilled and we just need to check whether (4.2) holds; otherwise, both conditions have to be examined. Unfortunately, the integrals in question are rather complicated and do not lend themselves to easy manipulation. Some more directly applicable results are grouped in the next two Theorems.

Theorem 7. *Let $0 \leq m < TN - k$. The m^{th} order marginal posterior moment of each of the components of β will exist if any of the following applies:*
(i) $p(z)$ is proper and the m^{th} order marginal prior moment of each of the components of z exists;
(ii) $p(z)$ is proper and $r(X : D) = k + M$;
(iii) $p(z)$ is bounded (not necessarily proper), $r(X : D) = k + M$ and $0 \leq m < TN - k - M$.

Proof: see Appendix. •

Finite prior moments of z will assure existence of posterior moments of β through Theorem 7 (i), whereas (ii) and (iii) exploit the structure of the model. However, this Theorem does not cover all situations of practical interest. Often the prior distribution for z is taken to be quite diffuse, so that the relevant prior moments may not exist. For example, the prior distributions for z used in the MIED and CED models of Koop, Osiewalski and Steel (1996) do not allow for first order marginal moments. In addition, as explained in discussing identification in Section 2, the design matrix $(X : D)$ is often of deficient column rank, which precludes the application of Theorem 7 (ii) and (iii).

In particular, whenever the frontier includes an intercept, β_1, and $D\iota_M = \iota_{TN}$, such as in Cases (ii), (iiia) and (iva), the rank of $(X : D)$ will typically be $k + M - 1$. Interest will then focus on $(\beta_2, \ldots, \beta_k)$, since economically meaningful quantities, such as elasticities, are functions of $(\beta_2, \ldots, \beta_k)$ and do not involve β_1. This practically most relevant situation is covered by the following Theorem:

Theorem 8. *If $r(X : D) = k + M - 1$, the frontier has an intercept, β_1, and $D\iota_M = \iota_{TN}$, the positive marginal posterior moments of $\beta_2, \ldots, \beta_k$ exist up to the order $TN - k$ (not including), provided $p(z)$ is proper.*

Proof: see Appendix. •

5. CONCLUDING REMARKS

In this paper we have examined the linear stochastic frontier model with composed error. In a Bayesian framework with improper prior distributions, we investigate whether posterior inference can be conducted, and if so, which posterior moments exist. We pay particular attention to the context of panel or longitudinal data.

According to the deterministic structure imposed on the vector of inefficiencies, we distinguish different cases, each with a particular parameterization of the inefficiencies in terms of a vector $z \in \mathcal{Z}$. We find that besides properties of the prior distribution of z, characterized by its density $p(z)$, (Theorem 2), it is the model structure that crucially affects the existence of the posterior distribution (Theorems 1 and 3) through the rank of the entire design matrix corresponding to both the frontier and the inefficiency terms. Whereas the importance of $p(z)$ was already stressed in Ritter (1993) in the particular context of our Example 1, the relevance of the model structure derives from the realization that we are essentially dealing with a Normal linear regression model. Whenever the location of the latter has the same dimension as the vector of observables, existence of the posterior is destroyed by the simple fact that the model is overparameterized. Thus, changing the functional form of the frontier[6] or using different priors on the frontier parameters does not solve the issue. In addition, note that probabilistic links between inefficiency terms [as introduced in Koop, Osiewalski and Steel (1996)] do not add to the structure. Only deterministic links that reduce the dimension of $\mathcal{Z}$ affect the existence of the posterior distribution.

Posterior moments of efficiencies and the precision of the measurement error are proven to be finite for many cases of practical interest (Theorems 4 and 5). The existence of posterior moments of the frontier parameters is the focus of Theorems 6-8. Theorems 6 and 7 collect some general results, that do not cover all models used in practice. Theorem 8 contains the finding of most practical relevance.

Our most striking results provide a warning to practitioners against the use of models with too little structure under the usual noninformative prior on precision. In particular, the "cross-section" specification where each firm has a specific inefficiency term for each time period [Case (i)] is shown to require more prior information than the "usual" noninformative prior density in (2.4) [Theorem 1 (ii) and Proposition 1].

In the light of the present findings, some results from previous papers have to be seriously qualified and reinterpreted. Formally, van den Broeck, Koop, Osiewalski and Steel (1994) and Koop, Steel and Osiewalski (1995) use model (2.3) – (2.4) on pure cross-section data with Case (i). The same fundamental problem occurs in Koop, Osiewalski and Steel (1994), which is based on a model with flexible functional form as in Proposition 2, again using Case (i). Thus, the posterior distribution does not exist in the theoretical model underlying these studies. Of course, in the computer implementation of the empirical analyses, there will implicitly be truncation due to computer limitations. In particular, truncating the precision away from zero (by machine precision) and truncating precision and frontier parameters in the tails while noting that these papers use a proper $p(z)$, one could argue that the prior in (2.4) effectively becomes proper, thus leading to a well-defined posterior distribution.

Whereas the latter argument can be used to reinterpret the empirical results obtained in those papers, it must be stressed that this leaves ample room for arbitrariness (it is not clear what the effect of computer truncation in any given run will be) and is thus not very satisfactory in general. It would be much preferable to formally adopt a different prior, such as the slightly more informative prior in (3.3). However, in this particular example[7] , the posterior distribution of the precision parameter, τ, turned out to be very concentrated at relatively low values: the posterior mass associated with values of τ over 20 is virtually zero in all cases. Thus, using (3.3) with $n_0 = 0$ and $a_0 = 10^{-6}$ instead of (2.4) will, most likely, not make any noticeable difference, in view of the fact that both priors are virtually indistinguishable[8] over the relevant range of τ. The results of these three previous papers are thus formally incorrect, but could empirically be reinterpreted in the context of a prior

[6] In this context, we mention specifications to accommodate a flexible functional form as in Koop, Osiewalski and Steel (1994) or making the frontier parameters a function of time as in Koop, Osiewalski and Steel (1995). Typically, such changes will not affect the dimension of the location vector.

[7] All three papers mentioned above use the same data, namely the cross-section of $N = 123$ electric-utility companies presented in Greene (1990, Appendix).

[8] The ratio between (3.3) and (2.4) monotonically decreases from 1 for $\tau = 0$ to 0.9998 for $\tau = 20$.

like (3.3).

The lack of existence of the posterior distribution can, however, be clearly noticed in other empirical contexts. Koop, Osiewalski and Steel (1995) consider an application to the growth of OECD countries based on a panel with $T = 10$ and $N = 17$. However, the panel structure is not exploited for the structure of the model, since the changes of efficiencies over time are of particular interest and thus there are no exact restrictions linking inefficiencies over time. This implies that Case (i) is used. In this model changing the prior from (3.3) to (2.4) has dramatic consequences for the inference on τ. With (2.4) the estimated posterior distribution of τ has a lot of mass attached to sets of very large values for τ: the 97.5^{th} percentile of τ is of the order 10^{13}, clearly indicative of the fundamental problem. Using prior (3.3) with $n_0 = 0$ and $a_0 = 10^{-6}$, leading to a well-defined posterior distribution, the 97.5^{th} percentile of the posterior distribution of τ is now around 2000. In addition, the Gibbs sampler used for the actual numerical integration converges much more rapidly, as it does not need to rely on "machine truncation" in this case. Note, however, that in this particular example, posterior inference on efficiencies and frontier parameters is not greatly affected by the use of the prior in (2.4) and the ensuing existence problem.

In Koop, Osiewalski and Steel (1996) a panel of U.S. hospitals was analysed with $N = 382$, $T = 5$, and $k = 38$. Here, the panel structure of the data was exploited in an "individual effects" model corresponding to Case (ii). The prior in (2.4) was used, but as $k < (T-1)N$ the proper prior distribution on z (for the MIED, VED and CED models) suffices to guarantee a well-defined posterior distribution [Theorem 1 (i)]. Posterior moments of efficiencies thus exist for all orders (Theorem 4) and for the scale parameter $\sigma = \tau^{-1}$ up to order $TN - k$, which is 1872 in this application (Theorem 5). As $r(X : D) = k + M - 1$ and the frontier includes an intercept, β_1, Theorem 8 assures us that positive posterior moments of the relevant frontier parameters, $\beta_2, \ldots, \beta_k$, also exist up to the order 1872.

Of course, the fact that in some empirical applications results were obtained that closely approximate their counterparts under a well-defined posterior, does not detract from the importance of the present results. Clearly, it is necessary to verify the existence of the posterior distribution and moments before conducting Bayesian inference. The present paper provides clear answers about these existence issues in the particular context of stochastic frontiers with composed error and suggests various ways of ensuring that a Bayesian analysis is feasible on theoretical grounds. Most importantly, if we have panel data at our disposal, we can make use of them in imposing some structure upon the model.

APPENDIX

Proof of Theorem 1

After integrating out τ in the Bayesian model in (2.3)−(2.4), existence of the posterior distribution is equivalent to the following integral being finite

$$\int_{\mathcal{B}\times\mathcal{Z}} \{(y - X\beta + Dz)'(y - X\beta + Dz)\}^{-TN/2} p(z) d\beta dz, \qquad (A.1)$$

for all $y \in \Re^{TN}$ except possible on a set of Lebesgue measure zero.

Part (i): $r(X:D) < TN$

Standard calculations show that

$$(y - X\beta + Dz)'(y - X\beta + Dz) = \{\beta - \hat{\beta}(z,y)\}'X'X\{\beta - \hat{\beta}(z,y)\} + c(z,y), \qquad (A.2)$$

where $\hat{\beta}(z,y) = (X'X)^{-1}X'(y + Dz)$ and

$$c(z,y) = (y + Dz)'M_X(y + Dz) = \{z + \hat{z}(y)\}'D'M_X D\{z + \hat{z}(y)\} + y'M_L y, \qquad (A.3)$$

with $M_X = I_{TN} - X(X'X)^{-1}X'$, $\hat{z}(y) = (D'M_X D)^{+}D'M_X y$, $L = (X : -D)$ and $M_L = I_{TN} - L(L'L)^{+}L'$; G^{+} denotes the Moore-Penrose inverse of a matrix G.

Clearly, $c(z,y) \geq y'M_L y$; furthermore, since $r(X:D) < TN$, we obtain $y'M_L y > 0$ barring a set of y's of Lebesgue measure zero. The integral in $(A.1)$ can now be rewritten as

$$\int_{\mathcal{Z}} \frac{p(z)}{\{c(z,y)\}^{(TN-k)/2}} \int_{\mathcal{B}} \{c(z,y)\}^{-k/2} \left[1 + \{\beta - \hat{\beta}(z,y)\}' \frac{X'X}{c(z,y)} \{\beta - \hat{\beta}(z,y)\}\right]^{-TN/2} d\beta dz. \qquad (A.4)$$

For the inside integral, we use the fact that the integrand is proportional to a k-variate Student-t density on β, whereas for the outside integral we use the bound $c(z,y) \geq y'M_L y > 0$. We then immediately see that $(A.4)$ is finite for any proper $p(z)$.

Part (ii): $r(X:D) = TN$

We show that all $y \in W$, with W as defined in (3.1), lead to an infinite integral in $(A.1)$.

Let $y \in W$; then $y = X\beta_0 - Dz_0$ for some $\beta_0 \in \mathcal{B}$ and $z_0 \in C$, and

$$(y - X\beta + Dz)'(y - X\beta + Dz) = (\delta - \delta_0)'L'L(\delta - \delta_0),$$

where $\delta' = (\beta', z')$, $\delta_0' = (\beta_0', z_0')$ and $L = (X : -D)$. By the Schur decomposition theorem $L'L = Q'\Lambda Q$, where Q is a $(k+M) \times (k+M)$ orthogonal matrix and $\Lambda = diag(\lambda_1, \ldots, \lambda_{k+M})$ is a diagonal matrix with the eigenvalues of $L'L$ as diagonal elements. Since $L'L$ is a positive semidefinite matrix of rank TN, it has TN non-zero eigenvalues, which will all be positive; without loss of generality we choose $\lambda_{TN+1} = \ldots = \lambda_{k+M} = 0$.

Since $z_0 \in C$ [defined in (3.2)], there exists a neighbourhood of z_0, say C', in which $p(z) \geq K > 0$ for some constant K. Therefore, a lower bound for $(A.1)$ is proportional to

$$\begin{aligned} \int_{\mathcal{B}\times C'} \{(\delta - \delta_0)'Q'\Lambda Q(\delta - \delta_0)\}^{-TN/2} d\delta &= \int_{\mathcal{N}} (\eta'\Lambda\eta)^{-TN/2} d\eta \\ &\geq \big(\max_{i=1,\ldots,TN} \lambda_i\big)^{-TN/2} \int_{\mathcal{N}} \Big(\sum_{i=1}^{TN} \eta_i^2\Big)^{-TN/2} d\eta, \end{aligned} \qquad (A.5)$$

for $\eta = (\eta_1, \ldots, \eta_{k+M})' = Q(\delta - \delta_0)$ and $\mathcal{N} = \{Q(\delta - \delta_0) : \delta \in \mathcal{B} \times C'\}$. We remark that the last integrand only involves the first TN components of η. Since the origin is an interior point of $\mathcal{N}$, we can find two balls $B(0, \varepsilon_1) \subset \Re^{TN}$ and $B(0, \varepsilon_2) \subset \Re^{k+M-TN}$, such that $B(0,\varepsilon_1) \times B(0,\varepsilon_2) \subset \mathcal{N}$. This implies that the last integral in $(A.5)$ has a lower bound proportional to

$$\int_{B(0,\varepsilon_1)} \Big(\sum_{i=1}^{TN} \eta_i^2\Big)^{-TN/2} d\eta_1 \ldots d\eta_{TN},$$

which, after a polar transformation, is immediately seen to be infinite.

Proof of Theorem 2

A lower bound for the integral in $(A.1)$ is obtained by substituting the entire domain of integration by $\mathcal{B} \times \mathcal{Z}_0$. Since $\mathcal{Z}_0$ is bounded, we can find a constant K such that $c(z, y) \leq K$ for all $z \in \mathcal{Z}_0$, for $c(z, y)$ defined in $(A.3)$. Applying the latter bound to $(A.2)$ leads to the following lower bound for the integral in $(A.1)$:

$$\int_{\mathcal{Z}_0} p(z) \int_{\mathcal{B}} [K + \{\beta - \hat{\beta}(z, y)\}' X' X \{\beta - \hat{\beta}(z, y)\}]^{-TN/2} d\beta dz.$$

Using again the fact that $\mathcal{Z}_0$ is bounded, we can assure the existence of a positive lower bound for the integral with respect to β, and we are left with the integral of $p(z)$ over $\mathcal{Z}_0$, which is infinite by assumption.

Proof of Proposition 2

The proof is similar to that of Theorem 1 (ii). After integrating out τ, we are left with

$$p(y) \propto \int_{\mathcal{B} \times \mathcal{Z} \times \Re^q} [\{y - X\beta + Dz - f(\alpha)\}' \{y - X\beta + Dz - f(\alpha)\}]^{-TN/2} p(\beta, z, \alpha) d\beta dz d\alpha. \quad (A.6)$$

We show that there exists a set of y's of positive Lebesgue measure in $\Re^{TN}$ for which the latter integral is infinite.

Choose $\alpha_0 \in \Re^q$ and $\varepsilon > 0$, such that the set

$$\begin{aligned} \mathcal{R} = \{(\beta, z) \in \mathcal{B} \times \mathcal{Z} : &\text{ for all } (\beta^*, z^*) \text{ in some neighbourhood of } (\beta, z) \text{ and all } \alpha \in B(\alpha_0, \varepsilon), \\ & p(\beta^*, z^*, \alpha) \geq K > 0 \text{ for some constant } K\} \end{aligned}$$

has positive Lebesgue measure in $\Re^{k+M}$.

Since $r(X : D) = TN$, the set $V = \{X\beta - Dz : (\beta, z) \in \mathcal{R}\}$ has positive Lebesgue measure in $\Re^{TN}$, and we can find a ball $B(v_0, \varepsilon_1) \subset V$ for some $\varepsilon_1 > 0$.

From the continuity of $f(\cdot)$, if ε is chosen to be small enough we obtain that for all $\alpha \in B(\alpha_0, \varepsilon)$, $f(\alpha)$ lies in the TN-dimensional ball $B(f(\alpha_0), \varepsilon_2)$ with $0 < \varepsilon_2 < \varepsilon_1/2$. It is then immediately obtained that

$$B(v_0 + f(\alpha_0), \varepsilon_2) \subset \bigcap_{\alpha \in B(\alpha_0, \varepsilon)} \{y \in \Re^{TN} : y - f(\alpha) \in B(v_0, \varepsilon_1)\},$$

and thus the set

$$W' = \bigcap_{\alpha \in B(\alpha_0, \varepsilon)} \{y \in \Re^{TN} : y - f(\alpha) = X\beta - Dz, \text{ for some } (\beta, z) \in \mathcal{R}\} \quad (A.7)$$

has positive Lebesgue measure. We now show that all $y \in W'$ lead to an infinite integral in $(A.6)$:

Using Fubini's theorem, we first integrate out β and z. For any $\alpha \in B(\alpha_0, \varepsilon)$ and $y \in W'$, we have that $y - f(\alpha) = X\beta_0 - Dz_0$ for some $(\beta_0, z_0) \in \mathcal{R}$, so that

$$\{y - X\beta + Dz - f(\alpha)\}'\{y - X\beta + Dz - f(\alpha)\} = (\delta - \delta_0)'L'L(\delta - \delta_0),$$

where $\delta' = (\beta', z')$, $\delta_0' = (\beta_0', z_0')$ and $L = (X : -D)$. The proof now proceeds similarly to that of Theorem 1 (ii) to show that for any $\alpha \in B(\alpha_0, \varepsilon)$ the inner integral with respect to β and z in $(A.6)$ is already infinite; this immediately implies that $(A.6)$ is also infinite for any $y \in W'$.

Proof of Theorem 6

After integrating out τ, $E(\beta_l^m|y)$ exists if and only if the following double integral is finite:

$$\int_{\mathcal{Z}} \frac{p(z)}{\{c(z,y)\}^{(TN-k)/2}} \int_{\Re^k} |\beta_l|^m \{c(z,y)\}^{-k/2} \left[1 + \{\beta - \hat{\beta}(z,y)\}' \frac{X'X}{c(z,y)} \{\beta - \hat{\beta}(z,y)\}\right]^{-TN/2} d\beta dz, \tag{A.8}$$

where we use the same notation as in the proof of Theorem 1 (i).

Part (i): $m \geq TN - k$

The inside integral in $(A.8)$ is proportional to an m^{th} marginal moment of a k-dimensional Student-t distribution with $TN - k$ degrees of freedom, which is infinite for $m \geq TN - k$.

Part (ii): $0 \leq m < TN - k$

Defining $\theta = \beta - \hat{\beta}(z,y)$, the inside integral in $(A.8)$ can be rewritten as

$$\int_{\Re^k} |\theta_l + \hat{\beta}(z,y)_l|^m \{c(z,y)\}^{-k/2} \left\{1 + \theta' \frac{X'X}{c(z,y)} \theta\right\}^{-TN/2} d\theta. \tag{A.9}$$

We now observe that

$$\max\{|\theta_l|^m, |\hat{\beta}(z,y)_l|^m\} I_{(0,\infty)}(\theta_l \hat{\beta}(z,y)_l) \leq |\theta_l + \hat{\beta}(z,y)_l|^m \leq 2^m \max\{|\theta_l|^m, |\hat{\beta}(z,y)_l|^m\}.$$

Applying these bounds to $(A.9)$ leads to an upper and lower bound for $(A.8)$ and Theorem 6 (ii) follows easily.

Proof of Theorem 7

For Parts (i) and (ii), (4.1) is clearly fulfilled since $p(z)$ is proper, so that we only need to verify that (4.2) holds.

Part (i): $p(z)$ proper with m^{th} order marginal moments.

From $(A.3)$ we know that $(y + Dz)'M_X(y + Dz) \geq y'M_L y > 0$, whereas

$$|w_l'(y+Dz)|^m \leq \left(|w_l'y| + \sum_{j=1}^{M} |w_l'd_j||z_j|\right)^m \leq (M+1)^m \left(|w_l'y|^m + \sum_{j=1}^{M} |w_l'd_j|^m |z_j|^m\right), \tag{A.10}$$

where $d_1, \ldots, d_M$ denote the columns of D. Applying the latter bounds to the integrand in (4.2), the result follows immediately.

<u>*Part (ii):*</u> $p(z)$ proper and $r(X:D)=k+M$.

We show that the product of the first two factors in the integrand in (4.2) is a bounded function of z, and thus the integral is finite for any proper $p(z)$. From $(A.10)$ with $m=2$ we deduce

$$|w_l'(y+Dz)|^2 \leq (M+1)^2\big(|w_l'y|^2+\|z\|^2\sum_{j=1}^{M}|w_l'd_j|^2\big),$$

where $\|\cdot\|$ denotes the Euclidean norm. Using both the latter bound and $(A.3)$, we obtain the following upper bound for the integrand in (4.2):

$$(M+1)^m(y'M_Ly)^{-(TN-k-m)/2}\{F(z)\}^{m/2}p(z),$$

where

$$F(z)=\frac{|w_l'y|^2+\|z\|^2\sum_{j=1}^{M}|w_l'd_j|^2}{y'M_Ly+\{z+\hat{z}(y)\}'D'M_XD\{z+\hat{z}(y)\}}.$$

We now show that $F(z)$ is a bounded function of z. Clearly, $F(z)$ is continuous, and therefore bounded in the region where $\|z\|\leq K$ for any positive constant K. We take $K>\max\{2\|\hat{z}(y)\|,|w_l'y|\}$, and we now examine the region where $\|z\|>K$: In this region

$$\frac{\|z\|}{2}\leq\|z\|\left(1-\frac{\|\hat{z}(y)\|}{\|z\|}\right)\leq\|z+\hat{z}(y)\|\leq\|z\|\left(1+\frac{\|\hat{z}(y)\|}{\|z\|}\right)\leq\frac{3\|z\|}{2}$$

and dividing both the numerator and the denominator of $F(z)$ by $\|z\|^2$ leads to

$$F(z)\leq\frac{1+\sum_{j=1}^{M}|w_l'd_j|^2}{\min\{\xi'D'M_XD\xi:\|\xi\|\in[1/2,3/2]\}},$$

where the denominator is bigger than zero since, due to the full column rank of $(X:D)$, $M_XD\xi=0$ only for $\xi=0$.

<u>*Part (iii):*</u> $p(z)$ bounded, $r(X:D)=k+M$ and $0\leq m<TN-k-M$.

Although it can also be shown that both (4.1) and (4.2) hold, this result follows immediately from the Normal linear regression model in (2.3) with the prior density $p(\beta,\tau,z)\propto\tau^{-1}$.

Proof of Theorem 8

We assume without loss of generality that $r(X:D_{M-1})=k+M-1$, where D_{M-1} denotes the first $M-1$ columns of D.

Consider the variable transformation from (β,z) to (δ,μ,z_M) such that $\delta=(\delta_1,\dots,\delta_k)'$ with $\delta_1=\beta_1-z_M$ and $\delta_l=\beta_l$, $l=2,\dots,k$, and where the $(M-1)$-dimensional vector μ groups the relative efficiencies z_l-z_M, $l=1,\dots,M-1$. Then

$$X\beta-Dz=X\delta-D_{M-1}\mu. \tag{A.11}$$

In the notation of Section 2, we have used $A=(I_M:-\iota_M)$ leading to $\tilde{\phi}_1'=(\delta_1,\mu')$ and $\phi_2=z_M$.

Furthermore, the transformation from (β, τ, z) to (δ, τ, μ, z_M) has unitary Jacobian and the prior in (2.4) leads to

$$p(\delta, \tau, \mu, z_M) \propto \tau^{-1} p(\mu, z_M),$$

where properness of $p(z)$ implies the same for $p(\mu, z_M)$.

Following $(A.11)$, z_M does not intervene in the likelihood and can be integrated out through $p(z_M|\mu)$. This leaves us with $k + M - 1$ regression coefficients, (δ, μ), and a full column-rank design matrix; thus, Theorem 7 (ii) applies and marginal posterior moments of $\delta_2 = \beta_2, \ldots, \delta_k = \beta_k$ exist up to the order $TN - k$.

REFERENCES

Aigner, D., C.A.K. Lovell, and P. Schmidt, 1977, Formulation and estimation of stochastic frontier production function models, Journal of Econometrics 6, 21-37.

Broeck, J. van den, G. Koop, J. Osiewalski, and M.F.J. Steel, 1994, Stochastic frontier models: A Bayesian perspective, Journal of Econometrics 61, 273-303.

Cornwell, C., P. Schmidt, and R.C. Sickles, 1990, Production frontiers with cross-sectional and time-series variation in efficiency levels, Journal of Econometrics 46, 185-200.

Fernández, C., J. Osiewalski, and M.F.J. Steel, 1995a, Inference robustness in multivariate models with a scale parameter, CORE Discussion Paper 9530. (Université Catholique de Louvain, Louvain-la-Neuve).

Fernández, C., J. Osiewalski, and M.F.J. Steel, 1995b, Modelling and inference with υ-spherical distributions, Journal of the American Statistical Association, 90, 1331-1340.

Geweke, J., 1995, Posterior simulators in econometrics, Research Department Working Paper 555. (Federal Reserve Bank, Minneapolis).

Greene, W.H., 1990, A Gamma-distributed stochastic frontier model, Journal of Econometrics 46, 141-163.

Jondrow, J., C.A.K. Lovell, I.S. Materov, and P. Schmidt, 1982, On the estimation of technical inefficiency in the stochastic frontier production function model, Journal of Econometrics 19, 233-238.

Koop, G., J. Osiewalski, and M.F.J. Steel, 1994, Bayesian efficiency analysis with a flexible form: The AIM cost function, Journal of Business and Economic Statistics 12, 339-346.

Koop, G., J. Osiewalski, and M.F.J. Steel, 1995, The components of output growth: A cross-country analysis, CORE Discussion Paper 9503. (Université Catholique de Louvain, Louvain-la-Neuve).

Koop, G., J. Osiewalski, and M.F.J. Steel, 1996, Bayesian efficiency analysis through individual effects: Hospital cost frontiers, Journal of Econometrics, forthcoming.

Koop, G., M.F.J. Steel, and J. Osiewalski, 1995, Posterior analysis of stochastic frontier models using Gibbs sampling, Computational Statistics, 10, 353-373.

Kumbhakar, S.B., 1990, Production frontiers, panel data, and time-varying technical inefficiency, Journal of Econometrics 46, 201-211.

Magnus, J.R. and H. Neudecker, 1988, Matrix differential calculus with applications to statistics and econometrics (Wiley, New York).

Meeusen, W. and J. van den Broeck, 1977, Efficiency estimation from Cobb-Douglas production functions with composed error, International Economic Review 8, 435-444.

Mouchart, M., 1976, A note on Bayes theorem, Statistica 36, 349-357.

O'Hagan, A., 1994, Bayesian Inference, Vol. 2B from Kendall's Advanced Theory of Statistics (Edward Arnold, London).

Osiewalski, J. and M.F.J. Steel, 1996, Numerical tools for the Bayesian analysis of stochastic frontier models, CentER Discussion Paper 9603. (Tilburg University, Tilburg).

Pitt, M.M. and L.F. Lee, 1981, The measurement and sources of technical inefficiency in the Indonesian weaving industry, Journal of Development Economics 9, 43-64.

Poirier, D., 1995, Revising beliefs in non-identified models, Mimeo. (University of Toronto, Toronto).

Ritter, C., 1993, The Normal-Gamma frontier model under a common vague prior does not produce a proper posterior, Mimeo. (Université Catholique de Louvain, Louvain-la-Neuve).

Schmidt, P. and R.C. Sickles, 1984, Production frontiers and panel data, Journal of Business and Economic Statistics 2, 367-374.

Stevenson, R.E., 1980, Likelihood functions for generalized stochastic frontier estimation, Journal of Econometrics 13, 57-66.

Zellner, A., 1971, An introduction to Bayesian inference in econometrics (Wiley, New York).